SOLUTION BRIEF

CLOUDAEON

Enterprise RAG Solutions
Built for Production

Stop hallucinations. Restore trust. Run RAG like a real system.

A production-grade RAG application deployed in your environment — fully owned by you, engineered for reliability, trust, and scale.

Explore the Problem →

cloudaeon.com/solutionpage/enterprise-rag

WHY ENTERPRISE RAG FAILS AFTER THE DEMO

Most RAG initiatives impress in demos — and collapse in production due to weak engineering.

TRUST KILLER

Hallucinations >30%

Inconsistent & unverifiable answers destroy user confidence at scale.

Root cause: Naive chunking + missing reranking / grounding validation

SYSTEM STALL

PoCs Stuck in Notebooks

Slow response times and rising costs as usage grows beyond demo.

Root cause: No AI Ops for quality, cost or drift management

GOVERNANCE VOID

No Evaluation Loop

Zero quality benchmarks, no ownership model, silent accuracy decay.

Root cause: Absence of continuous LLM-as-a-judge evaluation pipelines

SECURITY RISK

Access Control Gaps

Auditability and permission gaps block enterprise rollout entirely.

Root cause: No metadata-aware access control at the retrieval layer

See Our Solution →

ENTERPRISE RAG SOLUTION — CAPABILITIES

Metadata-Aware Ingestion

Preserves context, lineage & access controls across the pipeline.

Hybrid Retrieval + Reranking

Combines vector & keyword search to maximise relevance and recall.

Grounded Responses & Citations

Answer verification, traceability and user trust at scale.

LLM-as-a-Judge Evaluation

Built-in continuous, automated quality assessment pipelines.

Hallucination Detection

Measurable thresholds to monitor and control answer reliability.

Policy-Based Guardrails

Enforced consistently across the entire RAG lifecycle.

Secure RAG APIs

Controlled access with SSO/RBAC and optional UI.

CI/CD & Environment Promotion

Repeatable, controlled releases dev → staging → prod.

AI Ops Monitoring Dashboards

Tracks quality, latency and cost in real time.

Configurable Chunking & Embedding

Aligned to data types, document structure and use cases.

✦ **Delivered with perpetual license | Full source code handover | No vendor lock-in | No usage-based licensing**

RAG SOLUTION DELIVERY & COMMERCIAL MODEL

ONE-TIME IMPLEMENTATION

- ✓ Architecture finalisation aligned to your environment
- ✓ Deployment in the client environment
- ✓ System configuration & knowledge transfer

Core engagement — production readiness focused

OPTIONAL ONGOING SUPPORT

- ✓ SLA-backed AI Ops
- ✓ Evaluation tuning & optimisation
- ✓ Performance & cost optimisation
- ✓ New data source onboarding

Sustained operational assurance

OPTIONAL PROOF OF DESIGN

- ✓ Bespoke or regulated workflows
- ✓ Custom evaluation logic
- ✓ Agent or MCP integration
- ✓ Complexity de-risking — not mandatory

Used selectively for complex scenarios

INCLUDED ACCELERATORS

- RAG Evaluation Engine
- Guardrails & Safety Framework
- Metadata Ingestion Pipeline
- Doc Normalisation Engine
- Cost & Latency Playbooks

ENTERPRISE CONTRACT INTELLIGENCE PLATFORM

A large enterprise deployed Cloudaeon Enterprise RAG to power a contract intelligence platform operating at full production scale — with measurable, verified outcomes.

1,200+	<5%	97%	78%
Contracts Ingested	Hallucination Rate (down from ~28%)	Answer Accuracy via Continuous Eval	Effort Reduction in Contract Analysis

✦ **Transitioned from implementation to AI Ops within weeks, not months.**

KEY PRODUCTION KPIs TO TRACK

- Faithfulness
- Answer Relevancy
- Retrieval Precision @k
- Abstention Rate
- Latency p95
- Cost per Query
- User Adoption & Deflection Rate

If your RAG system isn't trusted, it isn't useful.

Take the first step with a structured, engineering-led approach.

Talk to a RAG Expert

cloudaeon.com/contact-us/solution-form

Registered: Mansion House, Manchester Rd, Altrincham, Cheshire, WA14 4RW, UK

© Cloudaeon Technologies Ltd. All Rights Reserved.